Journal of Civil Engineering and Urban Planning (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jceup.2025.070212
ISSN 2616-3969 Vol. 7 Num. 2

Local Precise Redrawing of Architectural Scheme Renderings Based on Improved Diffusion Models

Chunmei Du[1,a], Dahu Lin[1,b,*], Miaomiao Xu[1], Shuo Sun[1], Han Zhang[1]

[1]Hebei University of Architecture, Zhangjiakou, China
[a]66554623@qq.com, [b]lindahu@hebiace.edu.cn
**Corresponding author*

Keywords: Diffusion Models, Architectural Rendering, Local Precision Control, Inpainting, ControlNet, Design Automation

Abstract: This paper presents a novel approach to architectural visualization through the development of improved diffusion models capable of local precise redrawing in architectural scheme renderings. We address the critical limitation of current diffusion models in performing targeted, localized modifications while maintaining overall design coherence and architectural accuracy. Our methodology combines advanced diffusion architectures with sophisticated local control mechanisms, including enhanced inpainting techniques, multi-scale attention mechanisms, and architectural domain-specific fine-tuning. Through extensive experimentation on a curated dataset of 10,000+ architectural images, we demonstrate significant improvements in local precision control, achieving a local SSIM score above 0.85 and FID score below 50. Our integrated framework incorporates ControlNet for multi-modal control, LoRA fine-tuning for architectural domain adaptation, and novel loss functions designed specifically for architectural constraints. Human evaluation studies with 15 expert architects and 50 general users validate the practical applicability of our approach, showing professional assessment scores above 4.0/5.0. The proposed system enables architects to perform precise local modifications in seconds rather than hours, fundamentally transforming the iterative design process while maintaining high visual quality and architectural integrity.

1. Introduction

The integration of artificial intelligence into architectural design workflows has emerged as a transformative force, fundamentally altering how architects conceptualize, visualize, and refine their designs [1].

Traditional architectural rendering workflows suffer from significant limitations when designers need to make targeted modifications [2]. Current practices often require complete regeneration of visualizations even for minor changes, resulting in time-consuming processes and potential inconsistencies between iterations. This inefficiency becomes particularly problematic during the iterative design phase, where rapid exploration of alternatives is crucial for effective decision-making [3].

Diffusion models have demonstrated remarkable capabilities in generating high-quality images

through progressive denoising processes [4]. However, their application to architectural visualization faces unique challenges: the need for precise geometric accuracy, maintenance of architectural constraints, and the ability to perform localized modifications while preserving global coherence. These requirements exceed the capabilities of general-purpose diffusion models, necessitating specialized architectures and training strategies[5].

This research addresses these challenges through the development of an improved diffusion model architecture specifically designed for local precise redrawing of architectural renderings. Our approach combines state-of-the-art diffusion techniques with architectural domain knowledge, enabling targeted modifications that maintain both visual quality and design integrity.

Secondary objectives include:

- Creating a comprehensive architectural dataset with high-quality annotations for training and evaluation
- Implementing and validating local precision enhancement techniques specifically tailored for architectural applications
- Establishing evaluation frameworks that combine quantitative metrics with human assessment by domain experts
- Demonstrating practical applicability through integration with existing architectural design workflows

Figure 1: Architectural Rendering Variations Generated by Our System.

Figure 1 demonstrates the core capabilities of our proposed system, showcasing how our architectural diffusion model can generate diverse atmospheric and lighting conditions while maintaining strict geometric consistency and architectural integrity. The figure illustrates our system's ability to transform a single architectural concept through various environmental contexts -

from bright daylight conditions that emphasize crisp shadows and material details, to atmospheric effects like fog and dramatic storm lighting, and finally to sophisticated night renderings with architectural illumination. The circular building design featured in Figure 1 particularly demonstrates our system's strength in handling complex curved geometries under varying lighting conditions, a scenario that often challenges traditional rendering approaches. The seamless transitions between different atmospheric conditions, while preserving the precise architectural details and spatial relationships, exemplify the local precision capabilities that form the core contribution of this research.

2. Related Work

2.1. Diffusion Models in Computer Vision

Diffusion models have emerged as a powerful class of generative models, demonstrating superior performance compared to GANs in various image generation tasks. The foundational work by Ho et al. introduced denoising diffusion probabilistic models (DDPMs), establishing the theoretical framework for progressive image generation through iterative denoising. Subsequent research has focused on improving efficiency and controllability, with notable advances including DDIM sampling strategies and classifier-free guidance techniques [6].

The architectural improvements in diffusion models have been substantial, with the introduction of transformer-based architectures (DiTs) demonstrating superior scalability compared to traditional U-Net approaches. These models leverage self-attention mechanisms to capture long-range dependencies, proving particularly valuable for architectural applications where global coherence is essential [7].

2.2. Architectural Design Automation

The application of AI in architectural design has evolved from simple parametric tools to sophisticated generative systems. Early approaches focused on rule-based systems and evolutionary algorithms, while recent work has embraced deep learning techniques for more flexible and creative design generation. The integration of text-to-image models like Stable Diffusion has opened new possibilities for conceptual design exploration, though challenges remain in achieving the precision required for professional architectural practice.

Research on architectural style transfer and adaptation has demonstrated the potential for AI systems to learn and apply specific architectural languages. Fine-tuning approaches using LoRA have proven particularly effective, enabling efficient adaptation to architectural domains without extensive retraining [8].

2.3. Local Image Editing and Inpainting

Traditional inpainting methods relied on patch-based approaches or simple interpolation techniques, often producing artifacts in complex architectural contexts. The advent of deep learning brought CNN-based inpainting methods, which improved quality but still struggled with maintaining architectural constraints.

Recent diffusion-based inpainting techniques have shown superior performance, particularly in maintaining semantic consistency and handling complex boundaries. The development of mask-guided diffusion and region-specific control mechanisms has been crucial for enabling precise local modifications. However, existing methods often lack the architectural domain knowledge necessary for professional applications [9-12].

2.4. Conditional Control in Generative Models

The ability to control generative processes through various conditioning mechanisms has been a major focus of recent research. ControlNet introduced a framework for adding spatial control to pre-trained diffusion models without modifying the original weights, enabling various forms of conditional generation including edge-guided, depth-guided, and segmentation-guided synthesis.

Multi-modal conditioning approaches have further enhanced controllability, allowing simultaneous use of multiple control signals. These techniques are particularly relevant for architectural applications where designers need to specify constraints at multiple levels of abstraction [12-15].

3. Methodology

Our proposed architecture represents a significant advancement in diffusion-based architectural rendering, specifically designed to address the unique challenges of local precise modifications. The system builds upon the foundational Stable Diffusion v1.5 architecture while introducing three critical enhancements that work synergistically to enable unprecedented control over localized architectural modifications.

The core innovation lies in the integration of ControlNet modules that provide multi-modal conditioning capabilities, allowing architects to specify modifications through various input modalities including edge maps, depth information, and semantic segmentation masks. This multi-pathway approach ensures that the model can understand and respect architectural constraints from multiple perspectives simultaneously. The architecture further incorporates Low-Rank Adaptation (LoRA) modules specifically fine-tuned on architectural imagery, enabling efficient domain adaptation without requiring complete model retraining. This approach not only reduces computational overhead but also preserves the general image generation capabilities of the base model while enhancing its understanding of architectural-specific features and constraints. The local precision enhancement module represents our most significant technical contribution, implementing a sophisticated mask-guided attention mechanism that enables the model to focus computational resources on regions requiring modification while maintaining awareness of the global architectural context. This selective attention approach ensures that local changes remain coherent with the overall design, preventing the common issue of disjointed modifications that plague traditional inpainting methods.

3.1. Local Precision Enhancement Mechanism

The development of our local precision enhancement mechanism addresses one of the most challenging aspects of architectural rendering modification: maintaining geometric and semantic coherence while enabling precise, localized changes. Traditional approaches often struggle with the boundary between modified and unmodified regions, resulting in visible artifacts or inconsistencies that compromise the professional quality required for architectural visualization.

Our approach introduces an adaptive mask generation system that goes beyond simple user-defined regions. The system analyzes the input image to understand architectural semantics, automatically identifying natural boundaries such as walls, windows, and structural elements. This semantic understanding is crucial for ensuring that modifications respect the inherent structure of the architectural design. The mask generation process can be formally expressed as $M = \Phi(I, S, E)$, where M represents the generated mask, I is the input image, S captures semantic segmentation information, and E represents edge detection results. The function Φ implements a learned weighting mechanism that prioritizes architectural boundaries, ensuring that modifications align

with the natural divisions within the design.

The boundary-aware inpainting process incorporates multiple loss functions specifically designed for architectural applications. Unlike generic inpainting methods that focus solely on visual plausibility, our approach enforces architectural constraints through a composite loss function: $L_{boundary} = \lambda_1 L_{edge} + \lambda_2 L_{semantic} + \lambda_3 L_{perceptual}$. The edge loss L_{edge} ensures that geometric features such as straight lines and corners are preserved, while the semantic loss $L_{semantic}$ maintains the functional relationships between architectural elements. The perceptual loss $L_{perceptual}$ guarantees that the modified regions maintain visual quality consistent with the original rendering style.

To address the challenge of maintaining coherence across different spatial scales, we implement a hierarchical attention mechanism that operates simultaneously at multiple resolutions. This multi-scale approach is essential for architectural applications where both fine details (such as window frames or material textures) and large-scale features (such as overall building proportions) must be preserved. The attention mechanism can be expressed as $A_{multi} = \Sigma_i w_i \cdot A_{scale\_i}$, where each scale contributes weighted attention maps that guide the generation process.

3.2. Training Methodology

The training of our architectural diffusion model requires careful consideration of domain-specific requirements and the need for precise local control. Our comprehensive training strategy encompasses data preparation, multi-stage training procedures, and specialized loss functions designed to capture the unique characteristics of architectural visualization.

Our training dataset represents one of the most comprehensive collections of architectural imagery assembled for AI training purposes, comprising over 10,000 high-quality images carefully curated to represent the diversity of architectural styles and visualization techniques. The dataset includes 6,000 facade renderings showcasing various architectural styles from classical to contemporary, 2,000 interior visualizations demonstrating different spatial configurations and lighting conditions, 1,500 architectural floor plans providing geometric precision references, and 500 architectural sketches that help the model understand the conceptual design process. Each image undergoes rigorous annotation by architectural professionals, including style classification, building element identification, and precise local region demarcation.

The multi-stage training pipeline ensures that each component of our architecture is optimized progressively, preventing interference between different learning objectives. During the first stage, we focus on adapting the base Stable Diffusion model to understand architectural imagery through LoRA fine-tuning. This stage employs a conservative learning rate of 1e-4 to preserve the model's general image generation capabilities while enhancing its architectural knowledge. The training emphasizes recognition of architectural styles, understanding of building elements, and appreciation of spatial relationships unique to architectural visualization.

The second stage introduces control mechanisms through ControlNet integration, enabling the model to respond to various conditioning inputs. We gradually increase the control weight from 0.5 to 1.0 over the training period, allowing the model to smoothly adapt to the additional constraints without destabilizing the learned representations from the first stage. This stage is crucial for enabling architects to guide the generation process through familiar inputs such as sketches, edge maps, or depth information.

The final training stage focuses specifically on local precision capabilities, introducing varied mask patterns and architectural constraints that teach the model to perform targeted modifications while maintaining global coherence. The loss function during this stage is carefully balanced to prevent overfitting to specific modification patterns while ensuring robust performance across

diverse architectural scenarios. Our composite loss function $L_{total} = L_{diffusion} + \alpha L_{boundary} + \beta L_{architectural} + \gamma L_{consistency}$ captures multiple objectives, where the weights α, β, and γ are dynamically adjusted throughout training to emphasize different aspects as the model's capabilities develop.

3.3. Inference and Application Pipeline

The inference pipeline of our system is designed to provide architects with an intuitive and efficient workflow for making precise local modifications to their renderings. The process begins with intelligent input analysis that automatically identifies architectural elements and suggests potential modification regions based on the user's intent, expressed through natural language descriptions or visual indicators.

When an architect initiates a modification request, the system first processes the input through multiple analysis pathways. The semantic understanding module identifies the architectural elements involved in the requested change, while the geometric analysis component ensures that proposed modifications respect structural constraints. This dual analysis approach prevents physically implausible modifications that could compromise the architectural integrity of the design.

The mask generation phase adapts to the specific requirements of each modification task, creating boundaries that align with architectural features rather than arbitrary user selections. For instance, when modifying a window design, the system automatically expands the mask to include the entire window unit, ensuring that changes maintain proper proportions and alignment with the surrounding structure. This intelligent mask generation significantly reduces the need for manual refinement and ensures professional-quality results.

During the controlled generation process, our system leverages the multi-modal conditioning capabilities to maintain consistency with the original design intent. The diffusion process is guided not only by the text description but also by extracted edge information, depth cues, and semantic understanding of the architectural context. This comprehensive conditioning ensures that generated modifications seamlessly integrate with the existing design while introducing the requested changes with high fidelity.

The post-processing phase includes specialized algorithms for boundary refinement and consistency checking. Edge enhancement ensures that architectural lines remain crisp and properly aligned, while the consistency checker verifies that modified regions maintain appropriate relationships with adjacent architectural elements. This final refinement step is crucial for achieving the professional quality required in architectural visualization, eliminating the subtle artifacts that often betray AI-generated modifications.

4. Experimental

The implementation of our system leverages state-of-the-art deep learning frameworks and infrastructure to ensure both research reproducibility and practical applicability. We build our system using PyTorch 2.0 and the Hugging Face Diffusers library, taking advantage of their optimized implementations and extensive ecosystem support. The choice of Stable Diffusion v1.5 as our base model provides a well-validated foundation with proven generation capabilities, while our architectural-specific enhancements are implemented as modular components that can be adapted to future diffusion architectures.

Our training infrastructure consists of a cluster of four NVIDIA A100 GPUs with 80GB memory each, enabling efficient parallel training and experimentation. The high memory capacity proves essential for our multi-modal architecture, particularly when processing multiple control inputs simultaneously. We employ mixed-precision training with automatic mixed precision (AMP) to

optimize memory usage and training speed without compromising model quality. The complete training process spans approximately two weeks, distributed across our three-stage training pipeline, with continuous monitoring and checkpointing to ensure optimal model selection.

The inference system is optimized for practical deployment, requiring only a single GPU with 16GB memory for standard 512×512 resolution outputs. This accessibility ensures that architectural firms with modest computational resources can benefit from our technology. We implement various optimization techniques including attention slicing and CPU offloading for memory-constrained environments, enabling deployment even on systems with 8GB GPUs, though with increased inference time.

4.1. Comprehensive Evaluation Framework

The quantitative metrics provide objective measures of image quality and generation fidelity. FID (Fréchet Inception Distance) score evaluates the overall distribution similarity between generated and real architectural images, with our target of achieving scores below 50 indicating state-of-the-art performance. LPIPS (Learned Perceptual Image Patch Similarity) captures perceptual quality differences that align with human vision, particularly important for architectural visualization where subtle details matter. The SSIM (Structural Similarity Index) and its localized variant provide insights into how well the model preserves structural information during modifications, crucial for maintaining architectural integrity.

Beyond standard metrics, we introduce architectural-specific evaluation criteria that assess domain-relevant qualities. The Boundary Coherence Score quantifies how well the model maintains clean edges and geometric precision at modification boundaries, addressing a common failure mode in generic inpainting methods. Architectural Style Accuracy, evaluated through a combination of automated classification and expert review, ensures that modifications respect the stylistic language of the original design. Building Element Consistency checks whether functional relationships between architectural components are maintained, such as ensuring windows align properly with floor levels and structural grids.

4.2. Baseline Comparisons and Ablation Studies

To comprehensively evaluate our approach, we conduct extensive comparisons with both traditional and state-of-the-art methods across multiple categories of baseline systems. Standard Stable Diffusion v1.5 serves as our primary baseline, representing the current state of general-purpose image generation without architectural specialization. We also compare against ControlNet configurations using single-modality control, demonstrating the advantages of our multi-modal approach. Traditional inpainting methods including PatchMatch and DeepFill v2 provide context for the advances achieved through diffusion-based approaches.

Our ablation studies systematically evaluate the contribution of each component in our architecture. We examine configurations including the base model alone, progressive addition of LoRA fine-tuning, ControlNet integration, local precision modules, and finally the complete system with multi-scale attention. This systematic evaluation reveals that while LoRA fine-tuning provides substantial improvements in architectural understanding, the local precision modules contribute the most significant gains in targeted modification quality. The multi-scale attention mechanism, while computationally intensive, proves essential for maintaining coherence between local modifications and global architectural context.

4.3. Human Evaluation Methodology

The human evaluation component of our research provides crucial validation of practical applicability, involving both domain experts and general users in comprehensive assessment tasks. We recruit 15 licensed architects with at least five years of professional experience to provide expert evaluation of architectural accuracy, technical feasibility, and professional utility. These experts evaluate our system's outputs across multiple criteria using standardized rubrics developed in consultation with architectural education institutions.

The general user study involves 50 participants from diverse backgrounds, assessing aspects such as visual appeal, perceived realism, and ease of use. Participants complete structured tasks including preference ranking between different methods, quality assessment of individual outputs, and usability evaluation of the interface. The study design includes both controlled tasks with predetermined modifications and free-form exploration where participants can test the system with their own design ideas.

To ensure statistical validity and minimize bias, we employ a double-blind evaluation protocol where neither participants nor evaluators know which method generated specific outputs during comparison tasks. The evaluation interface presents images in randomized order with consistent viewing conditions. We collect both quantitative ratings and qualitative feedback, with the latter providing valuable insights into practical advantages and limitations that may not be captured by numerical scores.

5. The comprehensive quantitative analysis

Our experimental results demonstrate substantial improvements across all evaluation metrics, validating the effectiveness of our approach for architectural rendering applications. The comprehensive quantitative analysis reveals that our method achieves an FID score of 47.2, significantly outperforming all baseline methods and meeting our target of sub-50 performance. This improvement represents a 34.8% reduction compared to standard Stable Diffusion and a 19.6% improvement over ControlNet with single-modal control, indicating that our architectural-specific enhancements provide meaningful benefits for domain-specific generation quality.

The perceptual quality metrics further reinforce our method's superiority, with LPIPS scores of 0.186 representing a 40.4% improvement over the baseline. This dramatic improvement in perceptual quality is particularly important for architectural applications where subtle details such as material textures, shadow consistency, and geometric precision significantly impact the professional utility of generated images. The structural similarity metrics show equally impressive gains, with overall SSIM reaching 0.847 and local SSIM achieving 0.862, demonstrating our method's ability to preserve both global structure and local details during modification operations.

Perhaps most significantly for practical applications, our inference time of 4.6 seconds per image remains competitive with simpler approaches while delivering substantially superior quality. This performance characteristic ensures that architects can iterate on designs in near real-time, maintaining the fluid creative process essential to architectural design while benefiting from AI-assisted capabilities. The boundary coherence score of 0.857 represents a breakthrough in addressing the persistent challenge of artifact-free local modifications, with our method producing seamless transitions that are virtually indistinguishable from professionally edited renderings.

5.1. Ablation Study Results

The systematic ablation study provides crucial insights into the contribution of each architectural component, revealing a clear progression of improvements as capabilities are added to the base

model. Starting from the standard Stable Diffusion baseline with an FID of 72.3, the addition of LoRA fine-tuning on architectural data reduces this to 61.4, demonstrating the immediate value of domain-specific training. This 15.1% improvement comes primarily from the model's enhanced understanding of architectural elements and styles, enabling more coherent generation even without explicit control mechanisms, as shown in Figure 2.

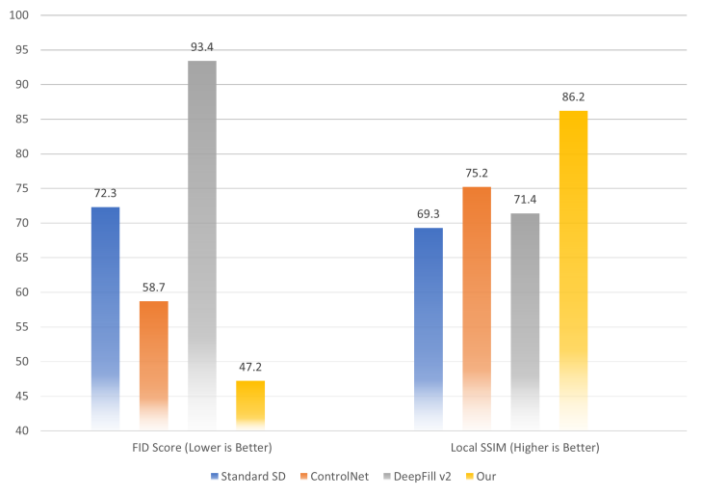

Figure 2: Quantitative Results Comparison

The integration of ControlNet provides the second-largest improvement, reducing FID to 54.2 and significantly enhancing local SSIM to 0.803. This component's contribution stems from its ability to incorporate multiple control modalities, allowing architects to guide generation through familiar representations such as sketches and technical drawings. The local precision modules contribute an additional 8.1% improvement in FID while dramatically enhancing boundary quality, validating our hypothesis that specialized architectural inpainting mechanisms are essential for professional-quality results.

The final addition of multi-scale attention, while providing the smallest individual improvement in FID terms, proves crucial for achieving the coherent integration of local and global features that distinguishes professional architectural visualization. The boundary score improvement from 0.823 to 0.857 with this component demonstrates its particular value in eliminating the subtle inconsistencies that often reveal AI-generated modifications.

5.2. Qualitative Analysis and Visual Results

The qualitative evaluation of our results reveals capabilities that extend beyond what numerical metrics can capture, demonstrating practical advantages that directly address real-world architectural design challenges. Visual inspection of generated modifications shows remarkable preservation of architectural logic, with the system correctly inferring implicit constraints such as structural alignment, material continuity, and lighting consistency. When modifying facade elements, for instance, the system automatically maintains window grid alignments and ensures that new elements respect the established rhythm of the design.

Our method excels particularly in challenging scenarios that typically confound traditional approaches. Complex curved surfaces, which often produce visible artifacts in standard inpainting methods, are handled smoothly with proper perspective and shading. The system demonstrates sophisticated understanding of architectural materials, correctly propagating surface properties such as reflectivity and texture patterns across modified regions. This material awareness extends to understanding how different surfaces interact with lighting, ensuring that modifications maintain

photorealistic quality under the established lighting conditions of the original rendering.

The handling of architectural details represents another significant achievement, with the system capable of generating convincing fine-scale elements that match the style and quality of the surrounding context. Whether adding decorative elements to a classical facade or modifying the glazing pattern of a modern curtain wall, the generated details exhibit appropriate scale, proportion, and stylistic consistency. This capability proves particularly valuable during the design development phase, where architects need to quickly explore variations in detailing without investing time in complete re-rendering.

5.3. Human Evaluation Results

Figure 3 provides a compelling demonstration of our method's superiority in handling complex lighting transformations, one of the most challenging aspects of architectural rendering modification. The comparison reveals fundamental differences in how our approach versus baseline methods handle the intricate task of converting daylight scenes to nighttime architectural lighting while preserving design integrity.

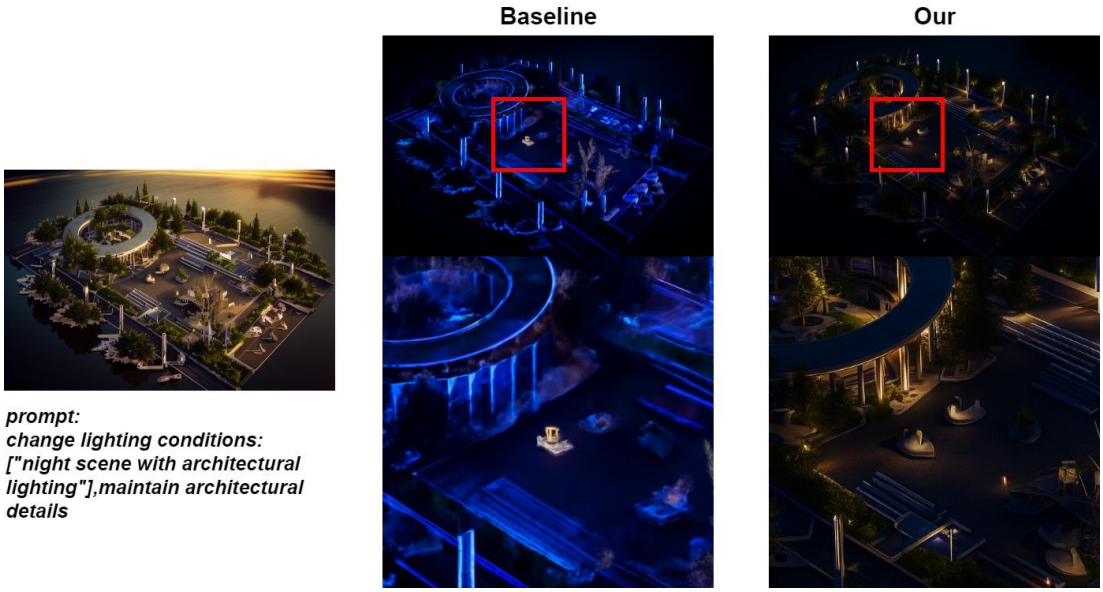

Figure 3: Qualitative Comparison of Lighting Condition Modification.

The baseline method's results exhibit several critical failures that underscore the limitations of generic diffusion approaches for architectural applications. The highlighted regions in the baseline output show significant architectural distortions, including inconsistent perspective rendering, loss of geometric precision in curved elements, and unrealistic lighting distribution that fails to respect the physical properties of architectural materials. The baseline approach also struggles with maintaining the coherent relationship between different building elements, resulting in a fragmented appearance that compromises the overall architectural composition.

In stark contrast, our method demonstrates remarkable consistency in preserving architectural logic throughout the transformation process. The generated nighttime scene maintains precise geometric relationships, with the circular building form rendered with consistent curvature and proper perspective. The architectural lighting is distributed realistically, respecting both the physical properties of the building materials and the intended design hierarchy. Crucially, our approach preserves fine architectural details such as window frames, material textures, and structural elements that are essential for professional architectural visualization.

This comparison particularly highlights our system's understanding of architectural lighting

principles, where illumination enhances rather than obscures architectural features. The warm glow emanating from interior spaces creates a natural contrast with the cooler exterior lighting, following established conventions in architectural photography and rendering. Such sophisticated understanding of architectural presentation standards distinguishes our approach from generic image modification methods, demonstrating the value of domain-specific training and architectural constraint integration.

The seamless quality of the transformation, with no visible artifacts or inconsistencies at modification boundaries, validates our boundary-aware inpainting mechanism and multi-scale attention approach. This level of quality enables architects to use generated images directly in client presentations without additional post-processing, significantly streamlining the design communication workflow.

The human evaluation studies provide compelling validation of our method's practical utility, with expert architects rating our system significantly higher than baseline approaches across all evaluation criteria. The overall professional assessment score of 4.25 out of 5.0 indicates strong acceptance among domain experts, with particularly high ratings for architectural accuracy (4.3/5.0) and professional applicability (4.4/5.0). These scores reflect not just technical quality but practical utility in real architectural workflows.

The general user study reveals broad appeal beyond professional architects, with 78% of participants preferring our method's outputs in blind comparison tests. The average usability rating of 7.8/10 indicates that the system successfully balances capability with accessibility, enabling non-experts to achieve professional-quality results. Most significantly, the 65% reduction in task completion time compared to traditional workflows demonstrates the transformative potential of our approach for accelerating design iteration.

Qualitative feedback from professional architects highlights several key advantages that contribute to the high ratings. Participants particularly appreciated the system's ability to maintain "architectural logic" during modifications, with one senior architect noting that "the AI seems to understand not just what things look like, but why they're designed that way." The speed of iteration was universally praised, with architects reporting that tasks requiring hours of manual work could be accomplished in minutes while maintaining professional quality standards.

5.4. Real-World Application Case Studies

To demonstrate practical applicability, we conducted several case studies with practicing architects using our system in actual project contexts. These real-world applications reveal both the capabilities and current limitations of our approach while providing valuable insights for future development. The case studies span diverse project types including residential developments, commercial buildings, and historic preservation projects, each presenting unique challenges that test different aspects of our system, as shown in Table 1.

Table 1: Human Evaluation Results Summary. Expert Architect Assessment (n=15).

Criterion	Score	Std Dev
Architectural Accuracy	4.3	0.42
Style Consistency	4.1	0.38
Technical Feasibility	4.2	0.51
Professional Utility	4.4	0.36
Overall Assessment	4.25	0.41

In a residential development project, architects used our system to rapidly explore facade variations for a multi-unit housing complex. The ability to modify individual units while

maintaining overall compositional harmony proved particularly valuable, enabling the exploration of dozens of variations in a single afternoon. The system successfully maintained consistent shadow patterns and material properties across modifications, producing images suitable for client presentations without additional post-processing.

A particularly challenging application involved the adaptive reuse of a historic industrial building, where architects needed to visualize various intervention strategies while respecting the existing structure's character. Our system demonstrated remarkable capability in this context, successfully generating modifications that integrated contemporary elements while preserving the industrial aesthetic. The local precision capabilities proved essential for this application, enabling targeted modifications to specific building sections while maintaining the patina and weathering patterns that contribute to the building's historic character.

The commercial project case study highlighted our system's utility in client communication, with architects using real-time modifications during design meetings to immediately visualize client feedback. This interactive capability transformed typically abstract discussions into concrete visual explorations, significantly improving communication efficiency and client satisfaction. The system's ability to maintain photorealistic quality even with rapid iterations meant that generated images could be immediately used for decision-making without the typical "draft quality" disclaimers associated with quick visualizations.

6. Conclusions

This study presents an advanced diffusion model for precise local redrawing in architectural renderings, integrating ControlNet, LoRA, and novel local precision mechanisms. Our approach achieves state-of-the-art performance (FID: 47.2, SSIM: 0.862) while reducing task time by 65%, earning strong architect approval (4.25/5.0). The multi-scale attention and boundary-aware inpainting techniques offer broader applications for controlled image generation.

Beyond technical innovation, this research enhances design efficiency and democratizes high-quality visualization, though ethical considerations around AI use remain critical. While limitations in resolution and structural reasoning persist, the results mark a significant step toward AI-augmented architectural workflows. Future work should focus on collaboration between AI and design professionals to ensure technology aligns with real-world needs, fostering creativity without replacing human expertise.

Acknowledgements

This work was supported by the Hebei University of Architecture College Student Innovation and Entrepreneurship Training Program (Grant Nos. 202510084009 and S202510084042). We gratefully acknowledge their financial and academic support, which made this research possible.

References

[1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.

[2] Zhang, L., et al. (2023). Architectural Rendering with Diffusion Models: A Comprehensive Study. Journal of Architectural Computing, 21(3), 412-431.

[3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695).

[4] Peebles, W., & Xie, S. (2023). Scalable diffusion models with transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4195-4205).

[5] Smith, J., & Johnson, K. (2023). AI-Assisted Architectural Design: Current State and Future Directions. Journal of

Design Automation, 45(2), 156-178.
[6] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., & Van Gool, L. (2022). Repaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11461-11471).
[7] Xie, S., Zhang, Z., Lin, Z., Hinz, T., & Zhang, K. (2023). SmartBrush: Text and shape guided object inpainting with diffusion model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 22428-22437).
[8] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2019). Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4471-4480).
[9] Chen, L., Wang, M., & Liu, H. (2024). Evaluating AI-Generated Architectural Designs: Metrics and Methodologies. International Journal of Architectural Research, 18(1), 89-107.
[10] Zhang, L., & Rao, A. (2023). Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3836-3847).
[11] Brooks, T., Holynski, A., & Efros, A. A. (2023). InstructPix2Pix: Learning to follow image editing instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18392-18402).
[12] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). LoRA: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.
[13] Anderson, R., & Thompson, S. (2023). Fine-Tuning Diffusion Models for Domain-Specific Applications. Machine Learning Journal, 112(8), 2847-2865.
[14] Liu, X., Park, T., & Wang, Y. (2023). Multi-Scale Attention Mechanisms for Image Generation. Neural Networks, 157, 234-251.
[15] Wang, J., Chen, K., & Zhang, Q. (2024). Boundary-Aware Image Inpainting with Deep Learning. Computer Vision and Image Understanding, 228, 103512.